Scholia Wikidata Graph Split Mitigation

Egon L. Willighagen[1], Wolfgang Fahl[2]

[1]Department of Translational Genomics, NUTRIM, Maastricht University, Netherlands
[2]BITPlan GmbH, Willich, Germany

Abstract
Wikidata was forced to perform a graph split in 2025 which will lead to the deactivation of the full legacy graph server in January 2026. The split has forced the Scholia project to overhaul the backend and query set being in use.

Keywords
Scholia, Wikidata

1. Introduction

The Wikidata graph split [1] has challenged the continuation of the Scholia service.

2. Background

Scholia [2] is a portal for the scientific community that provides access to Wikidata [3] curated content using a set of 387 queries organized in over 20 aspects. General aspects such as Author, Work, Venue, Topic, Organization, Publisher, Event and Location are covered as well as biochemical related ones such as Chemical compound, Gene/Protein, Disease and Taxon.

The Wikimedia Foundation provided Wikidata Query Service is using blazegraph as its SPARQL endpoint. The blazegraph 4TB storage limit is a threat to the continuation of the Wikidata service. Given the close to exponential growth of content since the launch in October 2012 the limit would be met within the upcoming months. Therefore the [1] graph split was announced and performed in 2025. The split led to having a wikidata-main graph with 8.6 billion triples and a wikidata-scholia graph with 8.7 billion triples instead of the full graph havin 17.1 billion triples.

3. Options

Given the deadline of January 2026 at which point the full graph provided by the Wikimedia Foundation will not be available any more the Scholia project was forced to take action and pursued the following options:

- modify affected queries to use federation accross the two split graphs
- migrate to a different backend such as QLever

4. Exploring Backend options

References

[1] Wikidata Community, Wikidata:SPARQL query service/WDQS graph split, Wikidata Documentation, 2024. URL: https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/WDQS_graph_split, accessed: 2025-12-06.

SWAT4LS 2026: Semantic Web Applications and Tools for Health Care and Life Sciences, 2026

[2] F. Nielsen, D. Mietchen, E. Willighagen, Scholia, Scientometrics and Wikidata, in: The Semantic Web: ESWC 2017 Satellite Events, ESWC 2017 Satellite Events, Springer International Publishing, Portorož, SI, 2017, pp. 237–259. URL: http://dx.doi.org/10.1007/978-3-319-70407-4_36. doi:10.1007/978-3-319-70407-4_36.

[3] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (2014) 78–85. URL: http://dx.doi.org/10.1145/2629489. doi:10.1145/2629489.